Understanding the Effects of Income on Housing Prices in the United States of America

Ryan Eastep
ECON 3161, Fall 2021
Georgia Institute of Technology
Dr. Shatakshee Dhongde
December 10, 2021

Abstract: This paper explores the effects of income on housing prices in the United States. Data from the 50 states including the District of Columbia was all collected within a year of each other in order to accurately reflect the metrics as they pertain to every part of the country. It is hypothesized that higher median income in a state would yield higher median housing prices, and for the most part we can conclude there is a significant correlation present. That being said, there are other factors which are shown to have as much if not a more significant impact on housing prices, such as household size, educational attainment, and statewide minimum wage.

Introduction:

In the modern-day United States, the balance between income and the prices of necessities in life has been one which has shown to be incredibly important for a functional society. The diversity in the nature of American households across the country as a result of conflicting cultures and environments illustrates its economic diversity, and thus it is of critical importance to understand how different income levels can affect the costs of the most important investments and purchases households need to make. In this paper, the primary focus will be how these changes in income reflect the costs of localized housing prices.

Housing is something all individuals in society must concern themselves with, due to being nothing short of an essential expense and a necessity for survival as well as a stable living situation. As a result, a proper balance between household income and housing prices is necessary to ensure the financial security of citizens. It is in the best interest of governments at the local and national level to keep people out of homelessness by providing affordable housing for those at various income levels. Furthermore, knowing the relationship between the two on a statistical level can provide important financial information for households and sellers operating in housing markets. Markets in theory require perfect information to reach maximum efficiency and understanding how income affects housing prices contributes significantly. Many statisticians have investigated the effects of income on housing prices, but it is rather rare to come across studies that have focused on this singular topic in favor of looking into other factors.

Of the studies that do exist concerning this topic, most conclude that income does yield an increase in housing prices. This paper will attempt to identify the direct impact of household income on US housing markets, controlling for external factors that are suspected to impact the housing market to a significant degree. As for a definitive hypothesis, we are expecting to observe that an increase in statewide median income should coincide with similarly increasing prices in statewide median housing prices. For this study, the primary independent variable will be the log of the median household income of a state, in order to fully realize the degree to which the economic circumstances differ across the country. The dependent variable will be the log of median housing prices per state, to similarly compare the independent variable directly to the housing prices local to each area.

Literature Review:

In a study researching impacts on regional housing prices, Dr. Alan K. Reichert (1990) focuses on the effects of interest rates, income, and unemployment. Pulling from US Census data, Reichert organizes the information collected into multiple tables detailing regional demographic information as well as economic data, making use of percent change in metrics such as population, real income, racial

diversity, as well as age and unemployment. These tables lead to correlation analyses investigating the relationship between real housing prices localized within nine major regions in the United States. Reichert ultimately discovers that housing prices in different regions react differently to shifts in specific trends. Furthermore, Reichert produces a results table showcasing the census data as it pertains to his metrics and how they affected housing prices in these major regions, with conclusions being made on the foundation on how each region performed relative to one another by not only just the raw numbers of change, but also the R^2 of each regression model. He concludes that, in America, the Northeast region is most reactive to population shifts, construction costs, and seasonality. In addition, the Middle Atlantic region is most reactive to employment rate. Mortgage rates most heavily influence New England housing prices, and permanent income has the largest demonstrable impact on the West. Dr. Reichert's findings help emphasize the notion of the US as that of a diverse economy, with regions responding differently to externalities. He also goes on to mention that the results of his research suggest that "enlightened housing policy and research should take into consideration both national factors as well as regional trends in income, employment, and key demographic characteristics" (Reichert, 1990, p. 388). The paper eventually transitions to his opinions against fully integrated national housing markets, as his data proves that regional differences cannot be ignored.

Dr. Viktorija Cohen (2017) takes a different approach to researching the effects on housing prices in her piece, entitled *The Analysis of the Determinants of Housing Prices*. Unlike Reichert's approach, which was rooted in demographic research as a core portion of the study, Cohen decides to ignore demographic information and focus entirely on economic parameters such as GDP, inflation, interest rate, and emigration. Cohen also decides to conduct her research based on data from Lithuania, as she mentions that developing countries are the area of interest for the study. Dr. Cohen sources her data mostly from Statistics Lithuania as well as the Bank of Lithuania, and cites that seasonality which may affect results had been removed by means of using the multiplicative method. Granger causality tests were conducted on the models provided, and eventually leads to the conclusion that "inflation, interest rate, and emigration are not casual determinants of average housing prices" (Cohen, 2017, p. 61). The implication in this study is that inflation cannot be utilized as an independent variable to interpret its effect on housing prices unless correlation and regression analyses had been adapted. However, significant relationships between GDP and unemployment and how they affect housing prices are present, with these variables being able to explain over 98% of housing price variation in Lithuania.

Lastly, in another study done outside the United States, Huiming Zhu, Zheng Li, and Peng Guo (2018) investigated determinants of housing prices, in their piece, *The Impact of Income, Economic Openness and Interest Rates on Housing Prices in China*. Considering how susceptible the determinants of housing prices can be depending on economic circumstances and general diversity in environment,

observing the effects on housing prices as they pertain to China specifically provides great insight into how these markets fluctuate across the world in comparison to the studies already reviewed in the US and Lithuania. The three authors base their research on data from 35 different major cities from 2002 to 2012 in mainland China, with the intention of utilizing dynamic panel quantile regression to obtain their results. Specifically, their data is pulled from China's National Bureau of Statistics, as the information is updated annually which is convenient for their intentions with the study. They arrive at the conclusion that the impact of income and population on housing prices in China is significant across all of their models, with economic openness following suit, although at a lower confidence level. (Zhu et. al, 2018, p. 4095). On the contrary, interest rates show some effect on housing prices, yet fall off drastically as more regressions were conducted. The study also concludes by discussing the implications of the study on public policy, specifically advocating for some form of government-controlled housing prices in certain cities, especially those where affordable housing is jeopardized as a result of the statistically significant effects of income.

This paper will research the same broader topic of determinants of housing prices, however it will focus primarily on just income in particular, as well as serve a different location of interest. While the above literature pieces research areas like the regional United States, Lithuania, and China this paper will focus primarily on the United States as a composition of 50 states and the District of Columbia to ultimately localize the data even further. Additionally, multiple regression models will be used incorporating income data alongside a unique set of other independent variables. These will include those that also affect housing prices, such as average household size, mortgage rates, vacancy rates, and more. Unlike the other studies, this paper will utilize natural logs of data consisting of high absolute values such as income and housing prices, the latter which will serve as the dependent variable for this study and the former being the primary independent variable. As for the datasets that will be used, rather than compiling time-series data, this study will focus on cross-sectional data collected either as close as possible to 2020, containing information on all 50 states and the District of Columbia, in order to test regression models against every part of the country rather than the US as a whole. In doing so, the effects of income on the affordable housing market can be examined without considering any noticeable changes that may have occurred over time. This ensures that the models will accurately incorporate the nuances of each state in order to gather a reliable conclusion on the dependent variable based on widespread data in the modern-day United States.

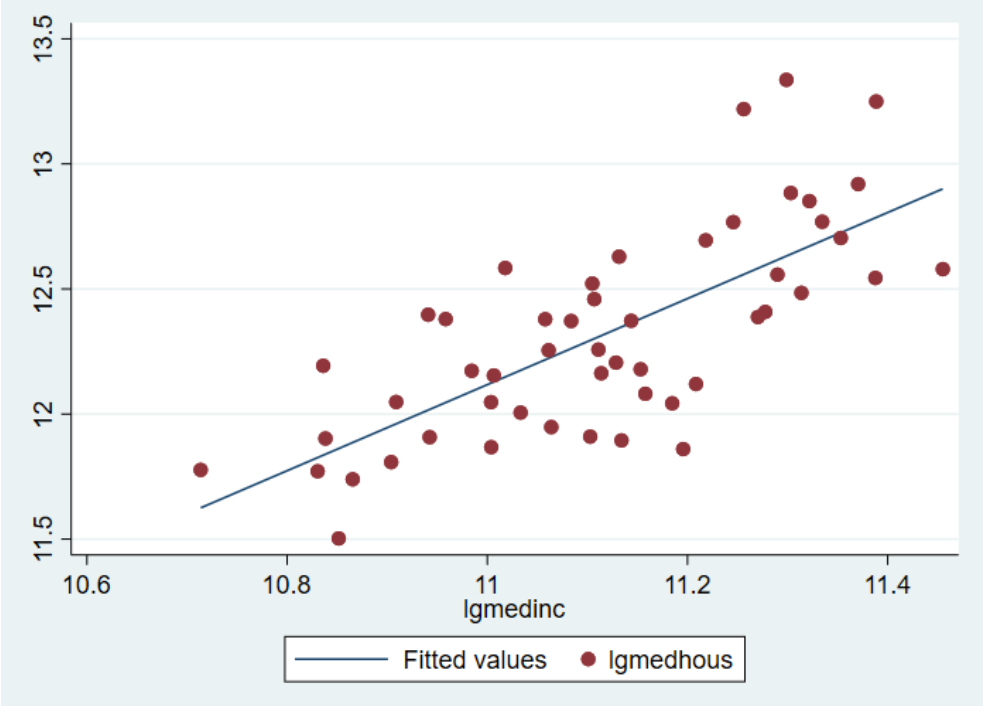
Data:

The data used in this study consists of information taken mostly from the US Census, the US Department of Labor, the US Department of commerce, third-party real estate companies which track

real-time trends in the online housing market such as Zillow and LendingTree. All of the data referenced was compiled from 2019-2021, as minimizing any externalities is of the utmost importance. It is also important to understand that some of the data, especially data collected during mid-late 2020, is at risk of being problematic due to drastic circumstances as a result of the COVID-19 pandemic, which can potentially affect the data in certain ways. However, this should not affect the modeling to such a degree that the results will be demonstrably different compared to data taken during another time period. Finally, as stated earlier, the data encompasses all 50 US in addition to the District of Columbia, which is fully detailed in Appendix A.

There are 9 variables total being considered: 1 dummy dependent variable, 1 dependent variable, and 7 independent variables. The main dependent variable of this study is *lgmedhous*, which is the natural log taken of each US state's median housing price. The primary independent variable, *lgmedinc*, represents the natural log of each US state's median income per household, which is the most effective metric for measuring income in its most standard form. As for the rest of the variables, other metrics including educational attainment, household size, vacancy rates, minimum wage, population, and mortgage rates are present. Below, represented in Figure 1 is a scatter plot showcasing the correlation between income and housing prices from the data, utilizing *lgmedhous* and *lgmedinc*. The data shows a clear positive trend with a decent amount of variance, but in encouraging moving forward regardless as it visibly supports the hypothesis of this study.

Figure 1: Income vs. Housing Prices



For brief context on the rest of the variables, *percbach* represents the general educational attainment of a state through the proportion of households with at least a bachelor's degree. The variable *housize* contains the average household size of each state, as it is hypothesized that larger household size would constitute the need for larger houses, which should ultimately correlate with an increase in housing prices. The variables *homvacan* and *renvacan* are home vacancy and rental vacancy rates respectively, and they were chosen to incorporate functions of consumer demand into the data. It seems logical that if vacancy rates are higher, it would imply a lack of demand for housing compared to other states with lower vacancy rates and thus decrease housing prices to accommodate the local housing market. The variable *avgmrgrate* tracks the average mortgage rate for each state, with the assumption that higher rates would discourage those living in the state from buying a house, which should imply higher housing prices. Lastly, the variable *coastal* is the lone dummy variable in this study, and simply tracks if a state is a coastal state or not. It was included under the assumption that coastal states generally have more expensive real estate to some degree by nature of being near the water, however it is unknown how much this affects housing costs so it was included for the sake of investigation. In Table I below, these variables and their descriptions are aggregated.

Table I: Description of Variables

Variable	Description	Year	Source
lgmedhous	Log of median housing price for states	2019	Zillow housing data & FICO score data from Experian
lgmedinc	Log of median income for states	2020	US Census
percbach	Proportion of households with at least a bachelor's degree for states	2020	US Census
housize	Average size of household for states	2020	US Census
homvacan	Homeowner vacancy rates for states	2020	US Census
renvacan	Rental vacancy rates for states	2020	US Census
avgmrgrate	Average mortgage rate for states	2019	LendingTree
minwage	Minimum wage for states	2021	US Department of Labor
lgpop	Log of state population	2021	US Census
coastal	= 1 if considered a coastal state	2021	US Department of Commerce

Furthermore, Table II below contains a numerical summary of the data researched for this study for all nine variables, derived from the STATA output shown in Appendix B. For the sake of simplicity, any data from this point forward will be rounded to two decimal places.

Table II: Numerical Summary of Variables

Variable	Observations	Mean	Std. Deviation	Min.	Max.
lgmedhous	51	12.32	0.41	11.50	13.34
lgmedinc	51	11.12	0.17	10.71	11.46
percbach	51	0.33	0.07	0.21	0.60
housize	51	2.54	0.17	2.28	3.08
homvacan	51	1.16	0.47	0.50	2.60
renvacan	51	7.07	2.81	2.50	16
avgmrgrate	51	4.85	0.05	4.74	4.98
minwage	51	9.58	2.30	7.25	15.20
lgpop	51	15.18	1.04	13.27	17.49
coastal	51	0.61	0.50	0	1

In anticipation of running regression models, the data was first checked to satisfy all Classical Linear Model (CLM) Assumptions:

1. **Linear in parameters:**
   → The models satisfy this assumption due to being linear in parameter upon the basis of the equation $y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \ldots + \beta_k x_k + u$

2. **Random Sampling:**
   → The data compiled is from all 50 states including the District of Columbia, without excluding any state or area in particular, thus satisfying this assumption.

3. **No Perfect Collinearity:**
   → None of the variables are perfectly collinear with any others, as seen in the correlation table in Appendix D. Additionally, none of the variables are scalar relative to another variable, evident by the fact that they are all fundamentally different categorically, so the assumption should hold.

4. **Expected value of the error term (*u*) equals 0:**

   → Based on the scatter plot detailing the main independent and dependent variables of the study, the line of best fit visible has roughly an equal concentration of data points both above and below it, failing to provide any clear evidence that the expected value of the error term would be anything apart from 0.

5. **Homoskedasticity:**

   → Also determined from observing the main scatter plot, homoskedasticity is clear in the consistent variance of the data points in regard to the line of best fit, allowing for it to be assumed for the rest of the study.

6. **Normality:**

   → Normal distribution will be assumed in order to allow for simple and multiple regression models to be calculated. Due to the presence of 51 observations for each variable this is an acceptable assumption considering the minimum requirement of 30 for normality.

Results:

Given that the models all have 50 degrees of freedom, the critical values measured are as follows as obtained from consulting the t-table: 1.299 for the 10% level of significance, 1.676 for the 5% level of significance, and 2.403 for the 1% level of significance. A majority of the variables are undergoing a one-sided test against the null hypothesis H_0: $\beta = 0$ and the alternative hypothesis H_a: $\beta > 0$, as they are hypothesized to have a positive impact on housing prices. The two exceptions are for tests involving vacancy rate related variables, which are testing with the alternative H_a: $\beta < 0$ due to the fact that their impact is hypothesized to be negative.

Simple Regression Model:

Model 1: lgmedhous = β_0 + β_1(lgmedinc) + *u*

Based on STATA output from Appendix C, the estimated equation for this regression model is:

Model 1: lgmedhous = - 6.81 + 1.72(lgmedinc)

n = 51 R^2 = 0.53

In this simple regression model, median income is the main and only independent variable being tested against the main dependent variable representing median housing prices. As visible in the STATA output, this model already produces an R^2 of 0.53 which is surprisingly high already. Considering this high value, it is not a stretch at all to imply that median income plays a rather significant role in determining housing prices. This is further supported by the high t-value of 7.45, which shows *lgmedinc* as statistically significant even past the 1% level. The R^2 tells us that upwards of 53% of the variance in housing prices in the data set can be explained by income, an encouraging sign that should be taken with a grain of salt nonetheless considering the nature of simple regression models. In order to fully understand the effects of income on housing prices, we need to incorporate the other control variables. Below Table III can be found, summarizing the results of this model.

Table III: Model 1 Estimation Results

Dependent Variable: lgmedhous			
Independent Variable:	Coefficient	Standard Error	t-statistic
lgmedinc	1.72	0.23	7.45

Multiple Regression Models:

Model 2: lgmedhous = β_0 + β_1(lgmedinc) + β_2(percbach) + β_3(housize) + β_4(homvacan) + β_5(renvacan) + β_6(avgmrgrate) + β_7(minwage) + β_8(lgpop) + β_9(coastal) + *u*

Based on the STATA output from Appendix D, the estimated equation for this regression model is:

lgmedhous = 13.51 + 0.089(lgmedinc) + 2.78(percbach) + 1.14(housize) + 0.00(homvacan) – 0.02(renvacan) – 0.95(avgmrgrate) + 0.04(minwage) – 0.11(lgpop) + 0.06(coastal)

n = 51 R^2 = 0.83

Being the first multiple regression model, this model incorporates all nine independent variables researched in the study to provide a starting point for further model building. The R^2 of 0.83 is the highest possible out of all the regressions in this study, due to including all variables, and thus proves itself to be the best model for predicting housing prices in this study. However, a few of these control variables, namely *homvacan*, *lgpop*, and *avgmrgrate* seem to raise issues upon further investigation. *renvacan* is significant at the 5% level and shows that rental vacancy rates, at least as shown in this model, have a larger impact on determining housing prices than housing vacancy rates. As for *homvacan,*

the t-statistic that result in this model for the variable is a pitiful 0.05, all but confirming it has a negligible impact on housing prices overall, and thus should probably be dropped for future models. In the case of *lgpop* in particular, it is quite surprising to see a negative coefficient, despite the earlier assumption that larger populations would yield higher housing prices due to inflated demand. This is likely a byproduct of outliers such as the District of Columbia having such astronomical housing prices despite its small population, and thus it will also be dropped for subsequent models. Additionally, *avgmrgrate* follows a similar trend, as it was expected that higher mortgage rates would result in higher housing prices, yet the coefficient that results is negative. It will be the third and final variable dropped for the creation of Model 3. As for *coastal*, its t-statistic is not significant even at the 10% level, however it is a rather interesting variable and will be kept regardless out of curiosity. Below Table IV can be found, summarizing the results of this model.

Table IV: Model 2 Estimation Results

Dependent Variable: lgmedhous			
Independent Variable:	Coefficient	Standard Error	t-statistic
lgmedinc	0.89	0.34	0.26
percbach	2.76	0.86	3.24
housize	1.14	0.19	5.91
homvacan	0.00	0.07	0.05
renvacan	-0.02	0.01	-1.94
avgmrgrate	-0.95	0.60	-1.58
minwage	0.04	0.02	2.57
lgpop	-0.11	0.03	-3.51
coastal	0.07	0.07	1.01

Model 3: lgmedhous = β_0 + β_1(lgmedinc) + β_2(percbach) + β_3(housize) + β_4(renvacan) + β_5(minwage) + β_6(coastal) + *u*

Based on the STATA output from Appendix E, the estimated equation for this regression model is:

lgmedhous = 3.82 + 0.44(lgmedinc) + 2.46(percbach) + 0.98(housize) – 0.02(renvacan) + 0.04(minwage) – 0.03(coastal)

n = 51 R^2 = 0.77

Model 3 builds upon the shortcoming of Model 2, specifically by retaining significant variables that produced visible impact. The relatively high R^2 value of 0.77 is quite impressive considering that the model was reduced in the number of variables, implying that it was the correct decision to keep the ones that were chosen to remain. At this point, it becomes clear which variables are the most critical to the model, as *percbach*, *housize*, and *minwage* all prove to be statistically significant even at the 1% level. The main independent variable, *lgmedinc*, regains some significance yet still seems to be lacking behind others. Moving forward, *coastal* and *renvacan* can be dropped to further focus on the more significant variables, as they simply do not provide the same level of insight the others do. Below Table V can be found, summarizing the results of this model.

Table V: Model 3 Estimation Results

Dependent Variable: lgmedhous			
Independent Variable:	Coefficient	Standard Error	t-statistic
lgmedinc	0.44	0.32	1.39
percbach	2.46	0.92	2.66
housize	0.98	0.20	4.95
renvacan	-0.02	0.01	-1.39
minwage	0.04	0.02	2.36
coastal	-0.03	0.07	-0.44

Model 4: lgmedhous = β_0 + β_1(lgmedinc) + β_2(percbach) + β_3(housize) + β_4(minwage) + *u*

Based on the STATA output from Appendix F, the estimated equation for this regression model is:

lgmedhous = 2.30 + 0.58(lgmedinc) + 2.20(percbach) + 0.95(housize) + 0.05(minwage)

n = 51 R^2 = 0.76

In this final model, the regression has been refined to the point where it becomes clear that the most significant determinants in explaining housing prices are still educational attainment, household size, and statewide minimum wage, similarly seen in Model 3. These three variables retain their significance including at the 1% level, further solidifying their status as objectively impactful. Circling back to the main variable of interest, income, it now becomes significantly impactful at the 5% level, all but confirming that income has a significant impact on housing prices when controlled for the proper variables. Below Table VI can be found, summarizing the results of this model.

Table VI: Model 4 Estimation Results

Dependent Variable: lgmedhous			
Independent Variable:	Coefficient	Standard Error	t-statistic
lgmedinc	0.58	0.30	1.89
percbach	2.20	0.91	2.42
housize	0.95	0.20	4.85
minwage	0.05	0.02	3.12

Now that all of the regression models have been completed, a comprehensive summary regarding the coefficients, standard error, and significance of each variable used in the models can be found below in Table VII.

Table VII: Estimation Results Summary

Dependent Variable: lgmedhous				
Independent Variable:	SLR	MLR1	MLR2	MLR3
lg(medinc)	1.72*** (0.23)	0.09 (0.34)	0.44* (0.32)	0.58** (0.30)
percbach		2.78*** (0.86)	2.46*** (0.85)	2.20*** (0.91)
housize		1.14*** (0.19)	0.98*** (0.20)	0.95*** (0.20)
homvacan		0.00 (0.07)		
renvacan		-0.02** (0.01)	-0.02* (0.01)	
avgmrgrate		-0.95 (0.60)		
minwage		0.04*** (0.02)	0.04** (0.02)	0.05*** (0.02)
lgpop		-0.11 (0.03)		
coastal		0.07 (0.07)	-0.03 (0.07)	
No. of obs.	51	51	51	51
R^2	0.53	0.83	0.77	0.76

Significance levels: 10%*, 5%**, 1%***

Extensions:

In observing the correlation table from the STATA output in Appendix G, the value that stands out is the potential collinearity between *percbach* and *lgmedinc*, which has about a rather high 0.79 value. This is likely due to the general pathway to high income in the US being tightly correlated to educational experience, as it is commonly known that those with a college degree tend to fare better financially in the labor market compared to those without. That being said, this leads to a necessary F-test for testing their joint significance as multicollinearity is an area of concern. For interpreting the calculated f-value, our null hypothesis is as such:

$$H_0: \beta_1 = \beta_2 = 0$$

The SSR value for the unrestricted model as shown by STATA in Appendix D is 1.46, while the SSR value for the restricted model is 3.63 as seen in Appendix H, giving the f-value calculation below:

$$F = \frac{(SSR_r - SSR_{ur}) / q}{SSR_{ur} / (n - k - 1)} = \frac{(3.63 - 1.46) / 2}{1.46 / (51 - 9 - 1)} = 30.47$$

Given that $df_1 = 2$ and $df_2 = 41$ in the regression models, and referring to the f-table, we can see that the critical value at the 1% significance level is 5.16. The calculated f-value of 30.47 clearly falls within the rejection region, and we can thus reject the null hypothesis and ultimately conclude that *percbach* and *lgmedinc* are, in fact, jointly significant at the 1% level, essentially confirming their joint significance.

Logs were used in the data in an attempt to minimize the difference in absolute value, as most of the data was relatively small in this regard. Specifically, median income, median housing price, and population data were subject to having their natural logs taken to achieve this. Below is an additional results table illustrating the issues encountered with standard error and coefficients when creating a similar model where logs are not used, using variables *medhous*, *medinc*, and *population*, labelled as "Model NoLog", from which the STATA output can be found in Appendix I.

Table VIII: Model NoLog Estimation Results

Dependent Variable: medhous			
Independent Variable:	Coefficient	Standard Error	t-statistic
medinc	0.89	1.84	0.26
percbach	736651.6	310410	2.37
housize	322178.1	71395.83	4.51
homvacan	5111.61	22975.9	0.22
renvacan	-2251.07	4123.12	-0.55
avgmrgrate	5349.89	209952.3	0.03
minwage	14725.64	5834.95	2.52
population	-0.0014	0.0016	-0.88
coastal	-1827.73	22516.35	-0.08

$n = 51$, $R^2 = 0.72$

It is important to understand that its purpose is not to be part of the regression models used in the study, but to serve as an example for why logs were taken for these variables. Immediately, it becomes clear that the absolute values of these terms becomes astronomically larger and provides unnecessary clutter and confusion while attempting to interpret these values. In comparison to the Model 2 results visible in Table IV and Appendix D, the use of logs undoubtedly proves to be the correct method for the sake of clarity, and general modeling overall considering the increase in R^2 as well.

A single dummy variable labeled *coastal* was used in this study, as it was determined that identifying coastal states could be important in helping determine housing prices. This variable has a value of 1 if the US state is considered to be a coastal state under the jurisdiction of government officials, and a value of 0 if the state is landlocked. Naturally, it proved to be of limited use in creating the regression models, never reaching even the 10% level of significance, but provided an interesting alternate variable regardless.

Conclusions:

Overall, it is clear that income plays an undeniable factor in explaining and predicting trends in housing prices in the United States. That being said, it remains far from the best explanatory variable, at least within the scope of the research done for this study. Factors such as educational attainment, household size, and minimum wage proved to have a much larger impact in the many models within which they were able to show great significance. Regardless, as it pertains to the initial questions raised in this study, we can firmly say that increases in income tend to yield higher localized housing prices, confirming our hypothesis. That being said, despite the data clearly showcasing this trend both in the models and the scatter plot, it remains to be seen that this impact of income on housing prices exists on a significant scale like some of the other control variables.

Referring back to the research of Dr. Reichert (1990) for a moment, it is important to understand the importance of these older studies holding up today. Even in an ever-evolving housing market which has dramatically expanded in the modern era of the internet, this study helps prove that determinants such as income and others outlined in Dr. Reichert's research are still relevant today and have stood the test of time. They prove just how difficult it can be to integrate states within the US into some sort of nationalized housing market – the difference in metrics like income as well as others create vastly different scenarios economically, and this wildly affects housing prices across the country. Different policies and regulations on the housing market exist for reasons like this, and this study further cements that notion. Ultimately, the effects of income on housing prices across US States is something that cannot be ignored and should be taken into account when predicting fluctuations in the housing market alongside other factors shown to do the same.

References

Cohen, V., & Karpavičiūtė, L. (2017). The analysis of the determinants of housing prices. Independent Journal of Management & Production, 8(1), 49–63. https://doi.org/10.14807/ijmp.v8i1.521

Department of Commerce, National Oceanic and Atmospheric Administration, National Ocean Service. (2021). Coastline of the United States. Retrieved from https://www.infoplease.com/us/geography/coastline-united-states

LendingTree. (2019). U.S. states ranked by average mortgage rate. Retrieved from https://www.housingwire.com/articles/48165-this-is-how-mortgage-rates-vary-by-state/

Reichert, A. K. (1990). The impact of interest rates, income, and employment upon regional housing prices. The Journal of Real Estate Finance and Economics, 3(4). https://doi.org/10.1007/bf00178859

US Census Bureau. (2019). Educational Attainment. Retrieved October 13, 2021, from https://data.census.gov/cedsci/table?q=educational%20attainment&g=0100000US_0400000US72&tid=ACSST1Y2019.S1501&tp=true&hidePreview=true

US Census Bureau. (2019). Owner/Renter (Householder) Characteristics. Retrieved October 13, 2021, from https://data.census.gov/cedsci/table?q=Owner%2FRenter%20%28Householder%29%20Characteristics&tid=ACSST1Y2019.S2503&hidePreview=false

US Census Bureau. (September 17, 2020). Average size of households in the United States in 2019, by state [Graph]. In Statista. Retrieved November 01, 2021, from https://www.statista.com/statistics/242265/average-size-of-us-households-by-state/

US Census Bureau. (December 22, 2020). Resident population of the U.S. in 2020, by state (including the District of Columbia) (in millions) [Graph]. In Statista. Retrieved November 02, 2021, from https://www.statista.com/statistics/183497/population-in-the-federal-states-of-the-us/

US Census Bureau. (August 10, 2021). Median household income in the United States in 2020, by state (in current U.S. dollars) [Graph]. In Statista. Retrieved October 13, 2021, from https://www.statista.com/statistics/233170/median-household-income-in-the-united-states-by-state/

U.S. Department of Labor, Bureau of Labor Statistics. (2021, Sep. 30). Consolidated Minimum Wage Table. Retrieved from https://www.dol.gov/agencies/whd/mw-consolidated

Zhu, H., Li, Z., & Guo, P. (2018). The impact of income, economic openness and interest rates on housing prices in China: Evidence from dynamic panel quantile regression. Applied Economics, 50(38), 4086–4098. https://doi.org/10.1080/00036846.2018.1441512

Appendix:

Appendix A: List of US states + District of Columbia

Alabama
Alaska
Arizona
Arkansas
California
Colorado
Connecticut
Delaware
District of Columbia
Florida
Georgia
Hawaii
Idaho
Illinois
Indiana
Iowa
Kansas
Kentucky
Louisiana
Maine
Maryland
Massachusetts
Michigan
Minnesota
Mississippi
Missouri
Montana
Nebraska
Nevada
New Hampshire
New Jersey
New Mexico
New York
North Carolina
North Dakota
Ohio
Oklahoma
Oregon
Pennsylvania
Rhode Island
South Carolina
South Dakota
Tennessee
Texas
Utah
Vermont
Virginia
Washington
West Virginia
Wisconsin
Wyoming

Appendix B: Data Summary

```
. sum lgmedhous lgmedinc percbach housize homvacan renvacan avgmrgrate minwage lgpop coastal
```

Variable	Obs	Mean	Std. dev.	Min	Max
lgmedhous	51	12.31891	.411641	11.50288	13.33586
lgmedinc	51	11.11708	.1743436	10.71366	11.45513
percbach	51	.3270392	.0654434	.211	.597
housize	51	2.54098	.1665804	2.28	3.08
homvacan	51	1.156863	.4704274	.5	2.6
renvacan	51	7.066667	2.818416	2.5	16
avgmrgrate	51	4.845098	.0539026	4.74	4.98
minwage	51	9.578039	2.29446	7.25	15.2
lgpop	51	15.18246	1.043428	13.26534	17.49278
coastal	51	.6078431	.4930895	0	1

Appendix C: Simple Linear Regression STATA Output

```
. regress lgmedhous lgmedinc
```

Source	SS	df	MS		
Model	4.49713482	1	4.49713482	Number of obs	= 51
Residual	3.9752818	49	.0811282	F(1, 49)	= 55.43
				Prob > F	= 0.0000
Total	8.47241662	50	.169448332	R-squared	= 0.5308
				Adj R-squared	= 0.5212
				Root MSE	= .28483

lgmedhous	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lgmedinc	1.720192	.2310441	7.45	0.000	1.255891	2.184493
_cons	-6.8046	2.568844	-2.65	0.011	-11.96689	-1.642313

Appendix D: Multiple Linear Regression #1 (Model 2) STATA Output

```
. regress lgmedhous lgmedinc percbach housize homvacan renvacan avgmrgrate minwage lgpop coastal
```

Source	SS	df	MS	Number of obs	=	51
				F(9, 41)	=	21.83
Model	7.00948037	9	.778831152	Prob > F	=	0.0000
Residual	1.46293625	41	.035681372	R-squared	=	0.8273
				Adj R-squared	=	0.7894
Total	8.47241662	50	.169448332	Root MSE	=	.1889

lgmedhous	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lgmedinc	.0893928	.3426662	0.26	0.795	-.6026356	.7814211
percbach	2.775979	.8560315	3.24	0.002	1.047188	4.504769
housize	1.143654	.1935222	5.91	0.000	.7528281	1.53448
homvacan	.003602	.0672689	0.05	0.958	-.1322502	.1394542
renvacan	-.022974	.0118218	-1.94	0.059	-.0468485	.0009005
avgmrgrate	-.9482303	.6007305	-1.58	0.122	-2.16143	.2649695
minwage	.0427941	.0166702	2.57	0.014	.009128	.0764602
lgpop	-.1119427	.0318649	-3.51	0.001	-.1762951	-.0475902
coastal	.0662219	.0658567	1.01	0.321	-.0667784	.1992221
_cons	13.51314	5.473867	2.47	0.018	2.458445	24.56784

Appendix E: Multiple Linear Regression #2 (Model 3) STATA Output

```
. regress lgmedhous lgmedinc percbach housize renvacan minwage coastal
```

Source	SS	df	MS	Number of obs	=	51
				F(6, 44)	=	24.38
Model	6.51347921	6	1.08557987	Prob > F	=	0.0000
Residual	1.95893741	44	.044521305	R-squared	=	0.7688
				Adj R-squared	=	0.7373
Total	8.47241662	50	.169448332	Root MSE	=	.211

lgmedhous	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lgmedinc	.4442283	.3188633	1.39	0.171	-.1983984	1.086855
percbach	2.456476	.9248904	2.66	0.011	.5924818	4.32047
housize	.9791013	.1976538	4.95	0.000	.5807563	1.377446
renvacan	-.0180789	.0130423	-1.39	0.173	-.0443639	.0082061
minwage	.0436492	.0184668	2.36	0.023	.0064319	.0808666
coastal	-.0286898	.0655374	-0.44	0.664	-.1607719	.1033922
_cons	3.816266	3.239312	1.18	0.245	-2.712137	10.34467

Appendix F: Multiple Linear Regression #3 (Model 4) STATA Output

```
. regress lgmedhous lgmedinc percbach housize minwage
```

Source	SS	df	MS	Number of obs	=	51
				F(4, 46)	=	35.80
Model	6.41234354	4	1.60308589	Prob > F	=	0.0000
Residual	2.06007308	46	.044784197	R-squared	=	0.7568
				Adj R-squared	=	0.7357
Total	8.47241662	50	.169448332	Root MSE	=	.21162

lgmedhous	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
lgmedinc	.5750096	.3044933	1.89	0.065	-.0379035	1.187923
percbach	2.195368	.9073852	2.42	0.020	.3688961	4.02184
housize	.9453571	.1950422	4.85	0.000	.5527574	1.337957
minwage	.0533202	.0170993	3.12	0.003	.0189011	.0877392
_cons	2.295672	3.024516	0.76	0.452	-3.792363	8.383707

Appendix G: Correlation Table for Independent Variables

```
. correlate lgmedhous lgmedinc percbach housize homvacan renvacan avgmrgrate minwage lgpop coastal
(obs=51)
```

	lgmedh~s	lgmedinc	percbach	housize	homvacan	renvacan	avgmrg~e	minwage	lgpop	coastal
lgmedhous	1.0000									
lgmedinc	0.7286	1.0000								
percbach	0.7003	0.7928	1.0000							
housize	0.3837	0.1598	-0.0872	1.0000						
homvacan	-0.1297	-0.3606	-0.1106	-0.1232	1.0000					
renvacan	-0.4967	-0.4651	-0.3986	-0.0228	0.1648	1.0000				
avgmrgrate	-0.3936	-0.4136	-0.1794	-0.3672	0.2715	0.2303	1.0000			
minwage	0.6333	0.4953	0.6446	-0.0248	0.0219	-0.5055	-0.1165	1.0000		
lgpop	-0.0492	-0.0046	-0.0198	0.3589	-0.1284	-0.0648	-0.2386	0.0306	1.0000	
coastal	0.2726	0.3076	0.3203	0.1557	0.0463	-0.0657	-0.1189	0.2226	0.3884	1.0000

Appendix H: F-Test Regression (Restricted Model) STATA Output

```
. regress lgmedhous percbach lgmedinc
```

Source	SS	df	MS	Number of obs	=	51
				F(2, 48)	=	31.99
Model	4.84062347	2	2.42031174	Prob > F	=	0.0000
Residual	3.63179315	48	.075662357	R-squared	=	0.5713
				Adj R-squared	=	0.5535
Total	8.47241662	50	.169448332	Root MSE	=	.27507

lgmedhous	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
percbach	2.077859	.9752136	2.13	0.038	.1170603	4.038657
lgmedinc	1.101859	.3660661	3.01	0.004	.3658335	1.837884
_cons	-.6100835	3.821886	-0.16	0.874	-8.294501	7.074334

Appendix I: Model "NoLog", A Version of Model 2 Without Log Variables, STATA Output

```
. regress medhous medinc percbach housize homvacan renvacan avgmrgrate minwage population coastal
```

Source	SS	df	MS	Number of obs	=	51
				F(9, 41)	=	11.83
Model	4.6521e+11	9	5.1690e+10	Prob > F	=	0.0000
Residual	1.7918e+11	41	4.3703e+09	R-squared	=	0.7219
				Adj R-squared	=	0.6609
Total	6.4439e+11	50	1.2888e+10	Root MSE	=	66108

medhous	Coefficient	Std. err.	t	P>|t|	[95% conf.	interval]
medinc	.894222	1.84174	0.49	0.630	-2.825248	4.613692
percbach	736651.6	310410	2.37	0.022	109765.8	1363537
housize	322178.1	71395.83	4.51	0.000	177991.3	466364.9
homvacan	5111.605	22975.9	0.22	0.825	-41289.16	51512.37
renvacan	-2251.065	4123.124	-0.55	0.588	-10577.88	6075.752
avgmrgrate	5349.888	209952.3	0.03	0.980	-418657.4	429357.1
minwage	14725.64	5834.953	2.52	0.016	2941.714	26509.57
population	-.0013964	.0015957	-0.88	0.387	-.0046189	.0018261
coastal	-1827.725	22516.35	-0.08	0.936	-47300.42	43644.97
_cons	-1022890	1081419	-0.95	0.350	-3206860	1161079